DOCUMENT RESUME

ED 383 207 FL 023 007

| | |
|---|---|
| AUTHOR | Heller, Anne; And Others |
| TITLE | A Comparison of Listening and Speaking Tests for Student Placement. |
| PUB DATE | 95 |
| NOTE | 16p.; For complete volume, see FL 023 004. |
| PUB TYPE | Reports - Research/Technical (143) -- Journal Articles (080) |
| JOURNAL CIT | Edinburgh Working Papers in Applied Linguistics; n6 p27-40 1995 |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | Comparative Analysis; *Dictation; *English (Second Language); Foreign Countries; Higher Education; Interviews; Language Research; Linguistic Theory; *Listening Comprehension Tests; Second Language Programs; *Student Placement; Summer Programs; Test Format; *Test Use; *Verbal Tests |

ABSTRACT

This study investigated the utility and effectiveness of a simple dictation test for placement of students in a university's summer English-as-a-Second-Language program. The dictation test was explored as an alternative to the traditional oral interview, which had shown some problems of interrater reliability and consistency, resulting in inappropriate placements. The study examined: (1) the relationship between scores on the two tests and student placements, to see whether the dictation test produced a better placement, and (2) the relationship between interview and dictation scores, to see whether the latter could be used in place of the former. The procedure was piloted with 14 native English speakers and 16 non-native speakers, then undertaken fully with 263 students. Results indicate that the dictation scores were marginally less effective than interview scores in placing students, and that the dictation cannot be effectively substituted for the interview. However, it is suggested that the dictation test be retained as an additional measure in case of a request for transfer from one class to another. The dictation is appended. (Contains 18 references.) (MSE)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 383 207

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Brian Parkinson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

# A Comparison of Listening and Speaking Tests for Student Placement

Anne Heller, Tony Lynch and Linda Wright (IALS)

FL023007



BEST COPY AVAILABLE

# A COMPARISON OF LISTENING AND SPEAKING TESTS FOR STUDENT PLACEMENT

**Anne Heller, Tony Lynch and Linda Wright (IALS)**

*Abstract*

*Oral interviews for placement purposes on entry to IALS summer courses appear to have presented problems of inter-rater reliability of assessment and consistency of questions asked, and the resulting transfers of 'mis-placed' students have engendered dissatisfaction amongst students and staff alike. This paper reports on the CLASP (Comparing Listening and Speaking for Placement) project, which aimed to investigate whether a simple dictation test might provide a more objective and reliable alternative to an oral interview. The results suggest that the dictation might be used to supplement the information provided by the interview, but not to replace it.*

## 1. Introduction

Student placement for IALS General English summer (GESUM) courses is based on a combination of students' scores on a cloze test and a five-minute interview with a teacher. For some years there has been increasing dissatisfaction amongst GESUM teachers with this method of placement; they have felt that students are mis-placed because of an interview score which has not satisfactorily predicted their level of performance in class. Over the first three or four days of each course it has been common for the GESUM course director to have to deal with transfers between classes within the GESUM course, as the (lack of) fluency of some students becomes evident in class work, or students perceive themselves to be in the wrong group. These inter-class transfers cause disruption to both students and staff.

In order to reduce the subjectivity of the oral interview, IALS runs a comprehensive pre-summer briefing of new teachers on the oral interview grading scheme. Despite this, wide variations in scoring appear to persist. The GESUM course director and assistant course director for 1993 (the first and third co-authors of this paper) discussed the possibility that the problems associated with the placement interview might be due to differences in perception of the scoring system on the part of temporary summer staff and permanent staff. However, informal examination of the oral interview sheets of students transferred within GESUM suggested that permanent and temporary staff were equally likely to award scores substantially higher or lower than the level indicated by subsequent classroom performance. The problem deserved further investigation and gave rise to the CLASP project.

The aim of the project was to design and trial a simple listening (dictation) test that might replace the present first-day GESUM student interview. There were two main issues for investigation: firstly, to measure the relationship between scores on the two tests and students' placement, to see whether dictation produced a better match with final placement (i.e. after any transfer); secondly, to examine the relationship between interview and dictation scores and assess whether dictation could be regarded as a proxy for oral communicative ability.

## 2. Background

Although there seems to be a common assumption in the language teaching profession that an L2 learner's skills in listening and speaking are closely related, the empirical evidence is relatively scant. There are numerous claims in the practical literature of the link between listening and speaking in general (e.g. Wong 1987) or between listening and the pronunciation component of speaking (Gilbert 1987). On the other hand, the evidence from research is that the relationship is not quite as strong or straightforward as teachers would probably expect. In those research studies that have correlated students' scores on tests of speaking and listening, the correlations reported have generally been in the range 0.5-0.6 (e.g. Lalande and Schweckendieck 1986, Criper and Davies 1988, Ferguson and White 1994).

Dictation may not be the most obvious type of listening test type to propose as an alternative measure for speaking ability, since it conventionally involves reproduction (rather than interpretation) and requires transfer from aural to written medium (cf. Chaudron 1985). The immediate local argument for using dictation for placement was several years' experience at IALS of using dictation as part of a battery of entry tests for pre-sessional English for Academic Purposes courses, where it has been found to give an adequate indication of students' oral skills for class placement. However, no formal comparison of scores has been carried out. Indeed, in the research literature we have located only one study comparing performances on dictation and speaking: Bacheller (1980) found a correlation of approximately 0.6 - similar to the figures reported in studies comparing other forms of listening and speaking assessment, mentioned above.

Weir (1990) provides a useful summary of research into dictation, though with no specific focus on its relationship with speaking. He cites Valette (1977) and Oller (1979) as among those who have argued for the value of dictation as a test of overall proficiency. To this we might add a recent study at IALS (Lynch 1994) which found that dictation was the strongest predictor among three language measures (the others being vocabulary and writing) for non-native postgraduates' academic success. On the other hand, there have been arguments against dictation: Heaton (1975) claimed that it demanded so many different skills that it allowed no firm measurement of any one; Alderson (1978) argued that dictation sampled relatively low-level language skills and recommended that dictation tests should be designed to challenge short-term memory and should also feature spoken-style texts rather than written-style.

In short, there seemed to us to be no clear theoretical reason for not using dictation for placement purposes, assuming the text and test design met these concerns, if it could be shown to provide a more reliable basis for student placement.

The decision to use dictation raises various issues: choice of text (invented or discovered), mode of delivery (live or recorded), and criteria for marking (verbatim reproduction or semantically acceptable version). Taylor (1983) favoured the use of what he called 'raw' dictation, i.e. the presentation of spoken text reflecting the patterns of natural speech rather than the careful written-to-be-read style of speech often adopted for the traditional type of L2 dictation. He recommended using a classroom teaching episode selected from a lesson, then re-recorded and spaced before presentation as a dictation test. Weir (1990) makes the general point that the content of any listening test, including dictation, should be appropriate to the students' learning situation - a point of particular relevance for a short course of the GESUM type.

As far as the marking of a dictation test is concerned, Taylor (1983) advocated ignoring spelling and punctuation, and reported a high - though unspecified - correlation between dictation marked on that basis and general proficiency in an end-of-course test. Bacheller (1980) proposed a 'Scale of Communicative Effectiveness' (SCE), intended to be a measure of the learner's ability to capture meaning in rendering the surface form of segments of text dictated between pauses. Weir (1990) refers to 'some evidence' that marking dictations for semantic appropriacy is more reliable than using exact word choice.

There is, too, the wider question of whether dictation is a 'fair' test, in the sense that it requires a reliance on memory that may put L2 listeners at a disadvantage. Two studies comparing dictation performances of

native and non-native speakers showed no significant difference in the range and type of errors made (Fishman 1980, Voss 1984). Fishman concluded that dictation as a test does not disadvantage non-native speakers unfairly.

Having consulted the various sources above we decided that the CLASP dictation should be based on a recording of a native speaker in a 'natural' context, that is, not pre-scripted for the purpose of the test. The chosen text would be divided into segments of increasing length (number of syllables), in order to increase the memory-load and therefore the difficulty of the segments, so as to increase the test's ability to discriminate between GESUM learners, whose level of proficiency varies from post-elementary to upper intermediate/advanced.

## 3. Pilot study

### 3.1 Materials

The text used in the dictation was an edited version of a radio interview dealing with the topic of tourism in western Scotland. We judged its content to be fairly typical of the information that GESUM students are exposed to during their Edinburgh course, particularly on the cultural visits and weekend tours that are organised for summer students.

To the dictation text itself was added a short introduction (on both the recording and the student's test script), designed to make the content more accessible. Students were also encouraged (see paragraph 3 of the instructions in Appendix A) to guess if they were uncertain about the precise wording of the original. The dictation text itself contained a total of 121 words. By providing the opening word or words of each segment, we created a target text of exactly 100 words in 11 segments of between 7 and 16 syllables.

The dictation text and marker's guidelines are shown in Appendix B. Misspelt words were not to be penalised if the word suggested that the student had understood, and syntactic or semantic alternatives would be accepted if they were appropriate in context. For examples, see Appendix B.

### 3.2 Method

The pre-pilot version of the dictation test was tried out in March 1994 with two groups of English teachers: 14 native speakers (IALS staff) and 16 non-native speakers (attending an ESP teacher development course). Scores in the native group were in the range 97-100 and those in the non-native group fell between 68 and 94. On an informal assessment, these scores suggested that the test was not unreasonably demanding, in the light of the natives' performance, and appeared to discriminate even among non-natives with relatively high English proficiency, such as those in the pre-pilot.

As a result of comments from the pre-pilot subjects, a number of text and format adjustments were made to the test sheet. The pilot study proper took place in April 1994. The subjects (n=38) were students entering a full-time general English programme in the April-June term. The dictation was presented as one part of a battery of placement tests, the others being a cloze test and an interview; in this respect the pilot study simulated the GESUM testing situation.

The cloze test comprises 147 items deleted from 13 short passages and the maximum time allowed for its completion is one hour. The interview is based on a *proforma* interview sheet, which the teacher fills in while talking to each student individually. A grade is given in the range 1-5 which, with possible intermediate grades such as 4+ and 3-, represents a 15-point scale.

### 3.3 Results

A summary of the statistics for the three tests is provided below. Cloze scores are the number of correct answers; Interview scores are converted to a number between 1 and 15; and Dictation scores can be read as a percentage, since the maximum score is 100.

Table 1. CLASP pilot study: Overall descriptive statistics (n=38)

| | range | sd | mean |
|---|---|---|---|
| Cloze | 2-116 | 26.92 | 60.03 |
| Interview | 1-15 | 3.31 | 8.74 |
| Dictation | 2-87 | 22.76 | 36.42 |

The Cloze and Interview were marked in the customary way by the General English course staff, working as a team in the case of the Cloze and alone in the case of the Interview. In order to establish the inter-rater reliability of the Dictation, the three researchers marked all 38 scripts blind, i.e. without consulting each other and without access to the students' other scores, on the basis of the agreed marking protocol (Appendix B).

The mean Dictation figure in Table 1 represents an average of the scores awarded by the three researchers. Comparison of the scores assigned to the Dictation scripts revealed a high level of inter-rater reliability, as shown in the Pearson figures in Table 2.

Table 2. CLASP pilot study: inter-rater correlations on Dictation scores

| | Marker 2 | Marker 3 | Average |
|---|---|---|---|
| Marker 1 | 0.993 | 0.994 | 0.998 |
| Marker 2 | - | 0.988 | 0.996 |
| Marker 3 | - | - | 0.997 |

In all cases, $p<0.0001$

From these figures it is clear that use of the marking protocol produced a very high degree of consistency in marks assigned.

To examine the primary issue for investigation - the degree of fit between Dictation performances and class placement - the students' scores on all three language measures were correlated with their final class placement, yielding the results shown in Table 3.

Table 3. CLASP pilot study: correlations between entry tests and class placement

| | Dictation | Interview | Cloze |
|---|---|---|---|
| placement | 0.752 | 0.807 | 0.895 |

$p<0.01$ in all cases



BEST COPY AVAILABLE

The high correlation between Cloze and placement (almost 0.9) accords with expectations: Cloze scores are used as the primary means of dividing the students into teaching groups, with the Interview grade used as secondary evidence to adjust up or down if an individual student has scored markedly higher or lower than others with a similar Cloze result.

On the other hand it was rather surprising to find that the Interview/placement correlation was quite so high (approximately 0.8), given that the CLASP project was motivated by the perception that Interview scores could be unreliable. Such a high figure might be taken as evidence that there was more consistency among interviewers than we had supposed. However, as we mentioned earlier, the problem giving rise to CLASP was with the assumed unreliability of the Interview conducted under GESUM conditions, i.e. by a larger number of teachers including those without previous experience of the first-day placement testing.

At this pilot stage, we were interested to note the initial evidence that the relationship between Dictation and placement was not significantly lower (at 0.752) than that for Interview/placement. This suggested that it was worth running the main study under GESUM conditions.

The second issue of particular interest was the statistical relationship between subjects' performances on the three placement test elements. Table 4 shows the Pearson results for individuals' scores.

Table 4. Cloze, Interview and Dictation correlations

| | Interview | Dictation ave. |
|---|---|---|
| Cloze | 0.827 | 0.840 |
| Interview | - | 0.814 |

p<0.01 in all cases

Although the Interview/Dictation correlation is markedly higher than the figures reported in the previous literature reviewed in section 2, there is no real difference in the relationship between any two of the three measures. In short, the 'triangle' represented by the Cloze, Interview and Dictation measures is equal-sided, at approximately 0.8.

## 4. Main study

### 4.1 Materials

We decided to use unaltered the text used in the pilot study and to retain the format agreed on after the pre-pilot study. The GESUM interview sheets are similar to those used for the April-June term course, but are designed to elicit more information. Interview grades are assigned on a nine-point scale (plus intermediates) as opposed to the five-point scale used in the pilot. The third part of the placement battery, a cloze test of 146 items, is also different from the one used in April. Despite the differences of form and content in the two cloze tests, we felt able to make broad comparisons between performances in the pilot and main studies, since an internal IALS report had established a correlation of 0.95 between the two tests.

### 4.2 Method

#### 4.2.1 Subjects

The data for the main study are scores on the three tests (Cloze, Interview and Dictation) of a total of 263 subjects tested at the second, third and fourth intake points in the GESUM programme in July-September 1994. For students who stayed for more than one course, only the original set of entry scores is included in the data.

#### 4.2.2 General procedure

The first-day GESUM routine is for students to take the Cloze test first, which lasts one hour, with a 10-minute break for the one-to-one interview 'inserted' at different points for different individual students. The Cloze tests are marked by a team of teachers following a strict marking guide. The Cloze scripts are then attached to the Interview sheet (showing the student's grade) and passed to the course director, who allocates students to one of seven levels. The Cloze score is used as the benchmark for placement; the Interview grade (together with information about age, mother tongue, previous learning experience, etc.) is used to make fine-tuning adjustments to produce classes that are as homogenous as possible in terms of their ability to use English. Students return to the test centre after lunch to receive details of their class (level and location).

#### 4.2.3 Dictation procedure

The three student intakes investigated in this study took the Dictation after having been assigned to their class. It had been our intention to administer it on the same day as the two placement tests, but practical difficulties arose with the first intake and we agreed to the GESUM Course Director's suggestion that we move the Dictation to a later day in the first week. So instead of a single first-day dictation session for all incoming students, the Dictation test was administered by each class's main tutor, who also marked their papers. Although this was not the planned testing configuration, we believe that the testing and marking procedure followed in the main study reflects the way in which the Dictation would be used if adopted. There were grounds for confidence that allowing the teachers to administer and mark the Dictation individually (without the opportunity to consult colleagues about marking queries) should not significantly reduce the reliability of the scores assigned, given the very high reliability figures established in the pilot study for teachers following the Dictation marking protocol in isolation.

#### 4.2.4 Data analysis

In autumn 1994 the papers for Dictation, Cloze and Interview were collated and any incomplete sets were removed from the data set. The data for each student were codified as an identifier, the three test scores, class level, age, gender, employment status and first language. Statistical treatment was carried out using SPSS-X.

### 4.3 Results

A summary of the statistics for the three tests is provided below. The Cloze score shows the number of correct answers (out of 146); Interview scores are converted to a number between 1 and 26; and Dictation scores out of a maximum of 100. It should be borne in mind that only the Dictation figures are directly comparable with those in the pilot study, since the content of the tests in the two studies was identical.

Table 5. CLASP main study: Overall descriptive statistics (n=263)

| | range | sd | mean |
|---|---|---|---|
| Cloze | 7-99 | 19.30 | 51.47 |
| Interview | 3-26 | 4.16 | 13.64 |
| Dictation | 0-78 | 16.78 | 30.91 |

When compared with the pilot study (Table 1), the Cloze and Dictation figures in Table 5 show that the overall proficiency of GESUM students was slightly weaker than that of the April-June students - some 15% lower on both measures. Although the Interview values appear to be higher, this is explained by the shift to the nine-point summer scale.

Again, we consider the pragmatic issue first: how does Dictation compare with Interview in terms of its degree of fit with overall class placement? It will be recalled that the original impetus for the CLASP project came from dissatisfaction felt by both staff and students that some interview grades were unreliable. In assessing whether Dictation is a better predictor, we will make the simplifying assumption that, once any first-week transfers had taken place, the students were in the class which they and their teachers found appropriate. Table 6 shows Pearson correlation figures for the three tests and the student's final class level.

Table 6. CLASP main study: correlations between entry tests and class placement

| | Dictation | Interview | Cloze |
|---|---|---|---|
| placement | .525 | .613 | .801 |

p<0.01 in all cases

These data indicate a slightly stronger relationship between Interview grade and final placement than between Dictation score and placement. The strongest correlation with class placement is, as in the pilot study, that of performance on the Cloze test.

On the issue of the inter-measure correlations, it will be recalled that the pilot study had found uniformly high values, at around 0.8, among Cloze, Dictation and Interview. We were particularly interested, for the purposes of the current project, in the relationship between GESUM students' performances on Dictation and Interview. Pearson correlation results are shown in Table 7.

Table 7. CLASP main study: Cloze, Interview and Dictation correlations

| | Cloze | Interview | Dictation |
|---|---|---|---|
| Cloze | - | 0.562 | 0.559 |
| Interview | | - | 0.576 |

p<0.01 for all values



BEST COPY AVAILABLE

The measured relationship between all three tests is substantially lower than in the pilot study: approximately 0.6, as opposed to 0.8. But, as in the pilot study, there is again no real difference in the relationship between any two of the three measures, although here the Dictation/Interview correlation is the closest, at 0.576. In the pilot study Dictation/Interview showed the lowest correlation, though at 0.814.

In sum, the evidence from both pilot and main studies is that the Dictation would be marginally less useful than Interview in providing additional information to Cloze scores when making decisions about placement. The inter-test correlations in both studies suggest that Dictation shares as much common ground with the Cloze as it does with Interview. In both cases the strength of correlations was weaker in the larger sample of the main study than in the pilot.

## 5. Discussion

There are a number of factors that could have resulted in the generally lower correlations in the main study, compared with the smaller-scale pilot. Firstly, it is quite possible that the different circumstances under which the GESUM students took the Dictation had a negative effect on some performances. In the pilot study, incoming students had taken the Dictation as part of the placement battery on the first day of their course. Since the GESUM students took Dictation as an in-class activity two or three days into their course, their motivation may well have been different. For example, there is evidence, in the form of almost blank scripts, that some students with relatively high Interview scores had underachieved; it could be that they took the Dictation less seriously than the placement tests.

Secondly, although we did our best to ensure that the GESUM tutors gave identical instructions to their classes (see Appendix C), we had no direct control over the way in which each tutor actually presented the purpose of the dictation to their students.

Thirdly, there was inevitable variation in the acoustic quality of rooms and of cassette players, which may have played a part in the lower correlation between Dictation and Interview, compared with results from the single dictation test session in the pilot study.

A further possibility is that the GESUM tutors' marking was influenced by the fact that, unlike the pilot study, they were marking their own students' scripts. This could have led to either over- or under-marking, or both, depending on the tutor's perception of the individual student.

The various potential factors we have mentioned are based on the observation that the Dictation/Interview relationship found in the GESUM study seems low, compared to the pilot. However, since the values reported in the research literature on listening and speaking (mentioned in section 2) are generally in the range 0.5-0.6, it is quite possible that the pilot study correlation of some 0.8 is atypically high. Are there any grounds for believing that - apart from its size - the pilot population was different from the subjects in the main study? Arguably, one relevant difference was that many of the 38 pilot study subjects were continuers from the previous IALS term and had been in Britain for at least three months prior to the April placement tests. That period of exposure to spoken English may have enabled those who had arrived with relatively weak listening comprehension to raise their 'listening fluency' to a point where any initial imbalance between spoken- and written-medium skills was minimised. In the case of GESUM, on the other hand, a much smaller proportion of students were continuers, and the lower correlations among Cloze, Interview and Dictation scores may reflect greater intra-personal differences in proficiency in different L2 skill components on arrival in Britain.

To sum up, the 0.6 correlation for Dictation/Interview established in the main study is close to the levels reported elsewhere for listening/speaking measures. Again, as in the pilot study, the triangular relationship between Dictation, Interview and Cloze is to all intents and purposes equilateral. We have no evidence that listening (as represented by our Dictation) can be regarded as a proxy for speaking (as represented by the GESUM Interview). Listening and speaking, as tested in the CLASP study, are as different from each other as they are from the text skills required to do well on the Cloze.

This study arose from a desire on the part of GESUM course directors to make placement more reliable, in order to reduce the number of students (and the amount of staff and student time in what are only three-week courses) involved in inter-class transfers. It may be that a resolution of the problem that motivated the study has emerged, quite independent of the CLASP project. In summer 1994 the number of GESUM transfers was markedly lower than in previous years. We believe that this can be ascribed to any or all of the following four factors. Firstly, the 1994 GESUM courses were held for the first time in the main IALS building, rather than spread over the two or three buildings we have used previously, which are several minutes' walk away; having a more compact and convenient course site seemed to create a more cohesive and settled atmosphere among GESUM participants. Secondly, a larger proportion of the teachers than in other years had previous experience of GESUM in general and of the placement interview in particular. This may have enabled them to make a more informed assessment of individual students' oral ability relative to the demands of the course. Thirdly, the time allowed for our main summer pre-course briefing was extended from two days to three, and one of the effects of this was to make the teaching staff generally more relaxed about the first-day testing arrangements. Finally, there was an overall decrease in GESUM student numbers compared with 1993, resulting in less pressure on staff during the period when placement was being decided.

The interplay of these real-world factors independent of our CLASP study is a salutary reminder that 'ecological' factors in real-life teaching may exert a more powerful influence than the variables one chooses to manipulate experimentally for theoretical reasons.

## 6. Conclusion

The two interrelated issues we set out to investigate in the CLASP project were the relationship between scores on listening and speaking tests, and the possibility of using a listening test instead of a speaking test for placement purposes. On the first of these, we found that Dictation scores would have been only marginally less effective overall than Interview scores in providing information to supplement Cloze scores when placing students in GESUM classes. On the second, our conclusion is negative: we found that the Dictation cannot be regarded as a proxy measure for the oral productive skills involved in the GESUM Interview.

Our recommendation is for no change: the GESUM Interview should be retained alongside the Cloze for placement purposes. On the available evidence - and probably for a complex mixture of reasons such as those we have discussed - placement based on Cloze and Interview scores was more successful in 1994 than in previous years, in the sense that there were fewer requests for transfer to another class.

However, given our finding that Dictation has approximately the same relationship with placement as Interview does, we suggest that the Dictation should be retained as a fall-back measure, for the course director to use in cases where a student may wish to move class against the advice of her/his current teacher. The supplementary information the Dictation provides on a student's likely performance in class may prove helpful as a more objective measure than the Interview grade - especially in cases where a students feels s/he has been assessed unfairly by the interviewer.

CLASP raises a number of further questions related to placement and mis-placement. In particular, we are aware that the reasons why a student might want to move to a different class are not necessarily linguistic at all; they may be cultural, cognitive or affective - among others. Even when the reasons are linguistic, in the sense of being related to the learner's assessment of her/his level in English in relation to others' in the class, they may relate to different areas of linguistic competence. Are there common patterns in requests for a move? For example, do students who ask for a transfer to a lower class tend to cite as their reason the fact that they are unable to understand the English of the teacher or fellow students? Do those who request a higher class want to move because they feel their level of fluency of production is superior to that of the other class members?

One possible avenue for future research would be to explore in depth the reasons why individual learners request (and teachers recommend) a transfer to another class. Are these reasons open to remedy - for example, by adjusting the grouping within the class so as to keep particular students apart, or to bring them together - or are they perceptions that cannot easily be changed over the period of a three-week course - e.g. differences in perceived language learning needs? A micro-level study on these and similar issues could be useful in assisting course directors to negotiate solutions with individual learners that satisfy all parties in what can be a difficult period of mutual adjustment at the start of a course. Research-based insights that might help reduce the stresses and strains of testing - and being tested - for placement would surely be beneficial for both staff and students in an intensive course such as GESUM.

## References


Alderson J.C. 1978. 'A study of the cloze procedure with native and non-native speakers of English'. Doctoral thesis, University of Edinburgh.

Bacheller F. 1980. 'Communicative effectiveness as predicted by judgments of the severity of learner errors in dictation' in Oller and Perkins (eds) 66-71.

Chaudron C. 1985. 'A method for examining the input / intake distinction' in S. Gass and C. Madden (eds) Input in Second Language Acquisition. Rowley, Mass.: Newbury House. 285-305.

Criper C. and A. Davies. 1988. Edinburgh ELTS Validation Project Report. British Council / University of Edinburgh.

Ferguson G. and E. White. 1994. A Predictive Validity Study of IELTS. University of Cambridge Local Examinations Syndicate / University of Edinburgh.

Fishman M. 1980. 'We all make the same mistakes: a comparative study of native and nonnative errors in taking dictation' in Oller and Perkins (eds) 187-94.

Gilbert J. 1987. 'Pronunciation and listening comprehension' in Morley (ed.) 33-39.

Heaton J. 1975. Writing English Language Tests. London: Longman.

Lalande J. and J. Schweckendieck. 1986. '*Zertifikat Deutsch als Fremdsprache* and the oral interview: a comparison of test scores'. Die Unterrichtspraxis 19: 16-26.

Lynch T. 1994. 'The University of Edinburgh Test of English at Matriculation: validation report'. Edinburgh Working Papers in Applied Linguistics 5: 66-77.

Morley J. (ed.) 1987. Current Perspectives on Pronunciation: Practices Anchored in Theory. Washington, DC: TESOL.

Oller J. 1979. Language Tests at School: a Pragmatic Approach. London: Longman.

Oller J. and K. Perkins (eds) 1980. Research in Language Testing. Rowley, Mass: Newbury House.

Taylor C. 1983. 'Testing proficiency with raw dictation'. RELC Journal 5/1: 50-51.

Valette R. 1977. Modern Language Testing. New York: Harcourt Brace Jovanovich.





BEST COPY AVAILABLE

Voss B. 1984. 'Perception of first-language and second-language texts: a comparative study'. Bielefelder Beiträge zur Sprachlehrforschung 13/2: 131-53.

Weir C. 1990. Communicative Language Testing. Hemel Hempstead: Prentice Hall.

Wong R. 1987. 'Learner variables and prepronunciation considerations in teaching pronunciation' in Morley (ed.) 13-28.

## Appendix A

**DICTATION** **NAME:** _____

You are going to hear about Torosay Castle on the Scottish island of Mull. When the present owner, Chris James, inherited Torosay he found it was in need of repair. So to pay the bills, he opened the castle to the public and hired it out for parties.

1st hearing: You will hear the whole text once. Listen, but do not write anything down.

2nd hearing: Now the text will be played in 11 short pieces. For each one, you are give the opening word of words. Try to write down exactly what is said. If you are not sure, guess. Don't worry about spelling or punctuation.

1. Several years

2. had provided

3. in coping

4. but

5. Just

6. Chris James

7. with

8. I think I'd known

9. I think

10. A

11. It does

Finally, you will hear the whole text again once for you to check what you have written. **TOTAL:** _____

14

## Appendix B

### DICTATION (CLASP) - Marking instructions

Each word correct counts as one point. There are a total of 100 missing words in the dictation, shown in italics in the transcript below.

Check the student's answer against the original. Count the number of words correct and write in the figure for each box/sentence on the right-hand side.

Add up the total and write it in at the foot of the student's script.

**Count as correct:**

(a) words in the original text;

(b) words not in the original but appropriate in context, e.g. syntactic alternatives (singular for plural) or semantic alternatives (*'taste'* for *'thirst'* in sentence 4)

(c) misspelt words that suggest the student has understood (e.g. *'rainfull'* for *'rainfall'* in sentence 3, but not *'air'* for *'heir'* in sentence 6)

Count as wrong:

(d) inappropriate words not in the original (e.g. *'a Shelby's house'* for *'a showpiece house'* in sentence 10, or *'rainfall of malt'*).

(e) word salad (correct words in jumbled order)

| | |
|---|---|
| 1. Several years ***spent working on aid projects in hot countries.*** | 8 |
| 2. had provided ***little in the way of experience*** | 6 |
| 3. in coping ***with sodden walls and the rainfall of Hull*** | 8 |
| 4. but ***did give a thirst for the challenges of regeneration*** | 9 |
| 5. Just ***as well as at the age of only thirty*** | 9 |
| 6. Chris James ***fell heir to the post of laird*** | 7 |
| 7. with ***several thousand acres of land attached*** | 6 |
| 8. I think I'd known ***for many years that I might come home to it sooner or later*** | 13 |
| 9. I think ***when it actually came to it I had quite a few sleepless nights*** | 13 |
| 10. A ***showpiece garden and a showpiece house but no museums these*** | 10 |
| 11. It does ***lend itself to a party - it's got a lovely atmosphere*** | 11 |

N.B. ***it's*** or ***it has*** counts as 2 words

**Possible total = 100**



BEST COPY AVAILABLE

## Appendix C

Introduction to CLASP for GESUM teachers

**To be read out by teachers before administering the dictation:**

*This morning I am going to give you a short test to measure your listening ability. This will give me/us extra information about your English to help us in our teaching. It is **not** a part of yesterday's placement test.*

*The test is a kind of dictation. It is about a Scottish topic, so it should be interesting and useful for you. I will give you this (SHOW ANSWER SHEET) piece of paper to help you. On it we give you the first few words of each section. You must write what you hear. DO NOT WORRY ABOUT SPELLING OR PUNCTUATION. I will play a tape with the dictation on it. You will hear all the instructions on the tape.*

*Read your answer sheet now.*

(Allow 2 minutes)

*Now listen and do exactly what the tape tells you.*

(Switch on tape)